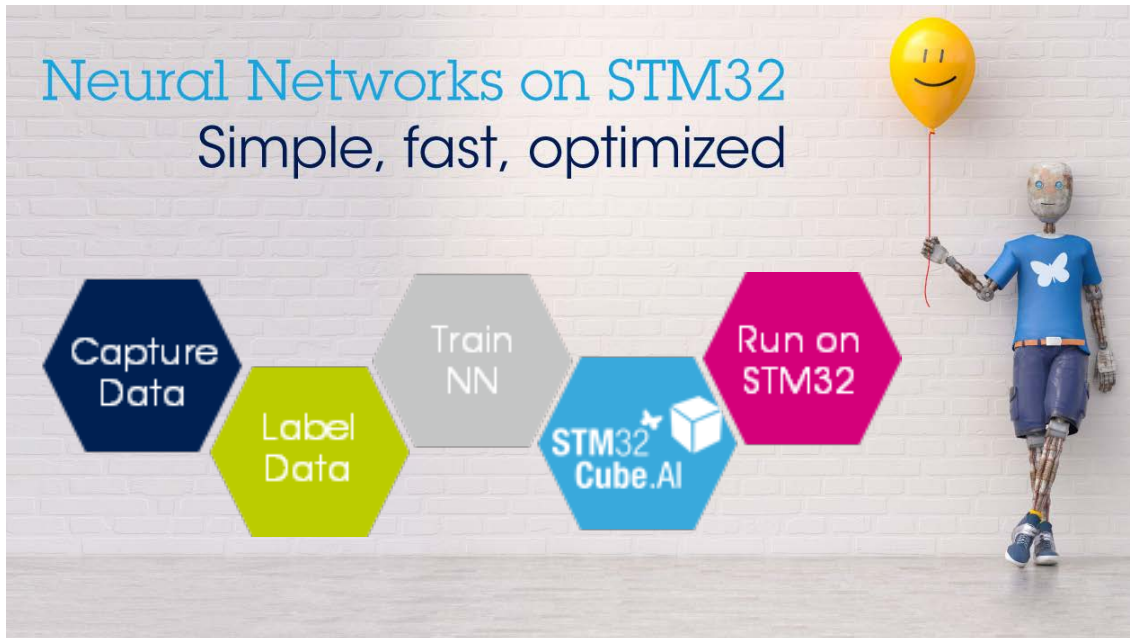


STM32Cube.AI – Neural Networks on STM32



ST Toolbox for Neural Networks

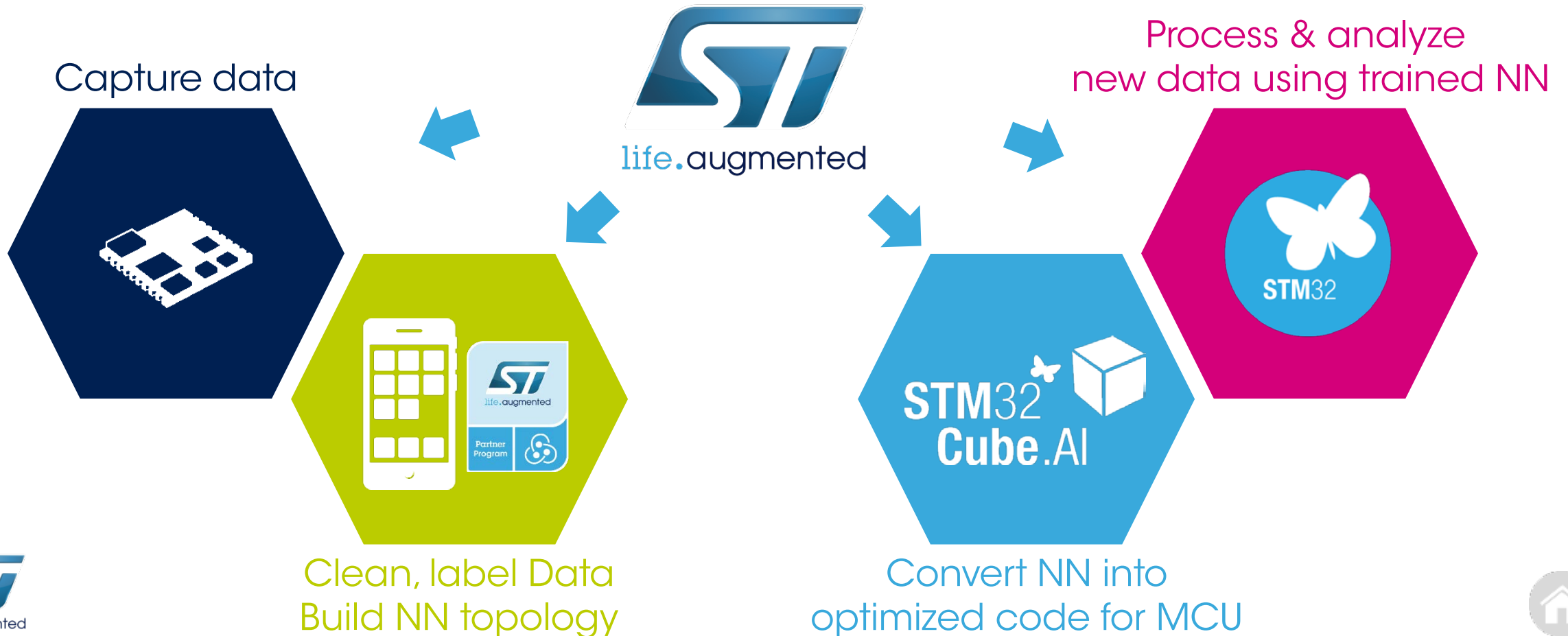
STM32Cube.AI

Form Factor HW to Capture and Process Data

Demos: Food Classification & Handwriting Character Recognition

ST Toolbox for Neural Networks

More Than Just a NN Conversion Tool



STM32CubeMX AI Extension

Brings AI To the STM32 MCU Family

Input your Framework dependent, Pre-Trained neural network into **STM32Cube.AI**

Automatic and fast generation of an STM32-optimized library

STM32Cube.AI guarantees interoperability with state-of-the-art Deep Learning Design Frameworks

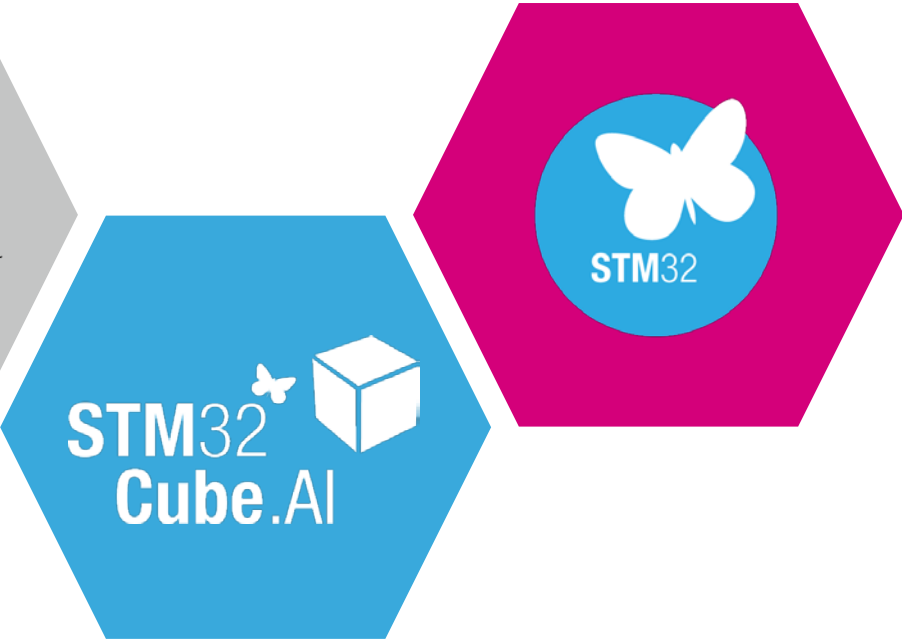
Train NN Model

Diagram showing supported frameworks for training: Keras (K), TensorFlow (T), ConvNetJS (Deep Learning in your browser), Lasagne, and Caffe.

Diagram showing frameworks supported soon: ONNX, Caffe2, Chainer, mxnet, PYTORCH, and CNTK.

Supported soon

Process & analyze new data using trained NN



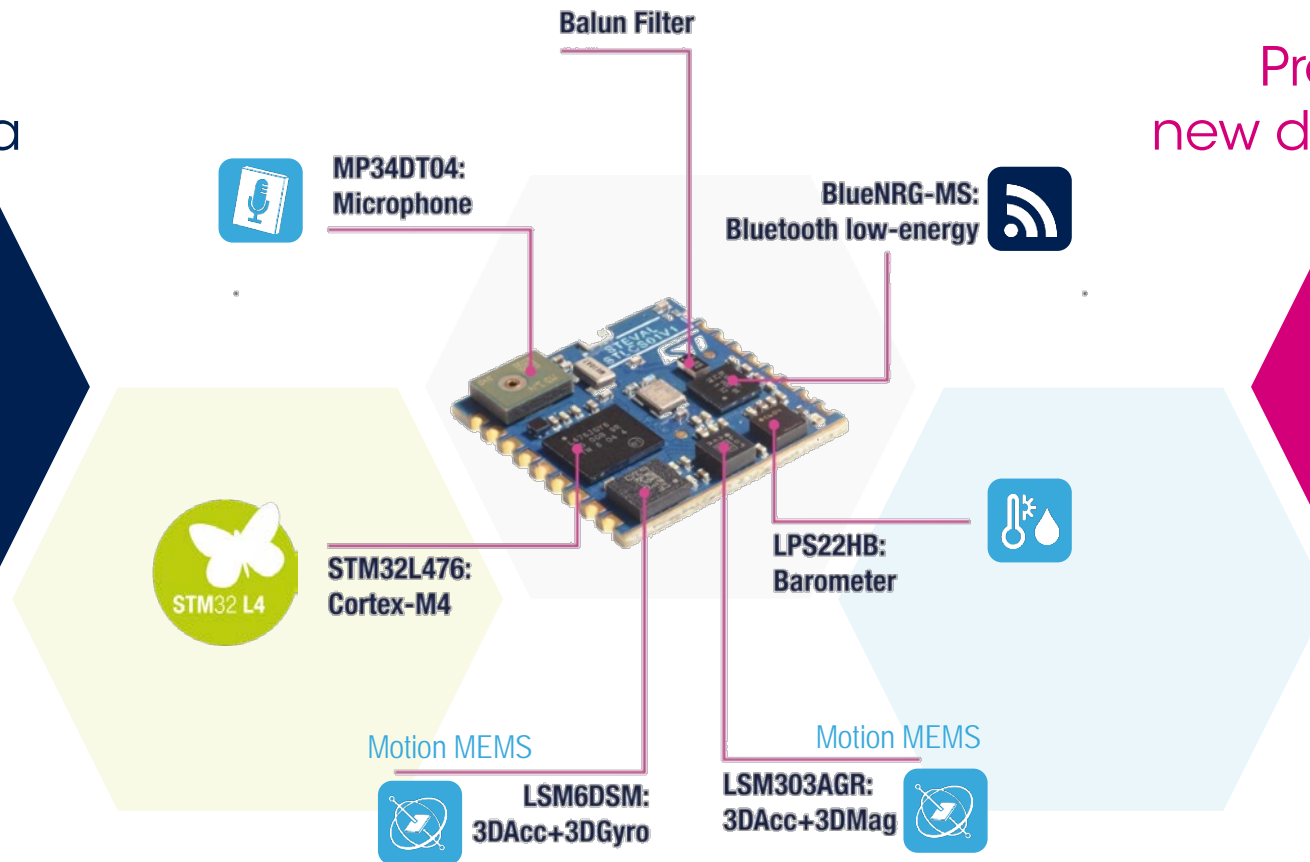
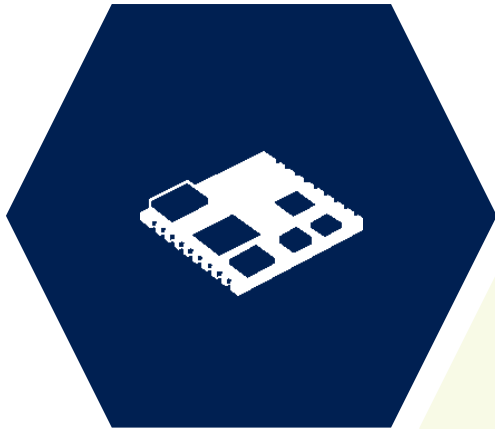
Convert NN into optimized code for MCU



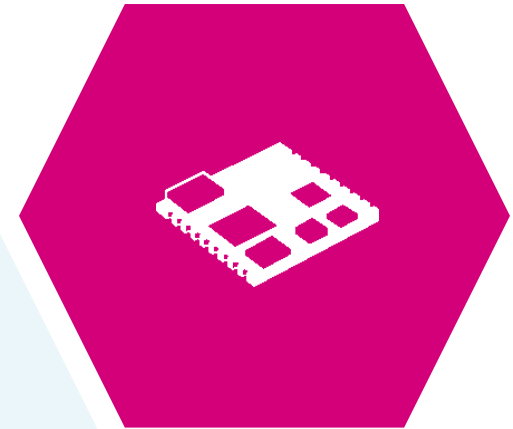
Form Factor HW

Capture and Process Data With the SensorTile

Capture data



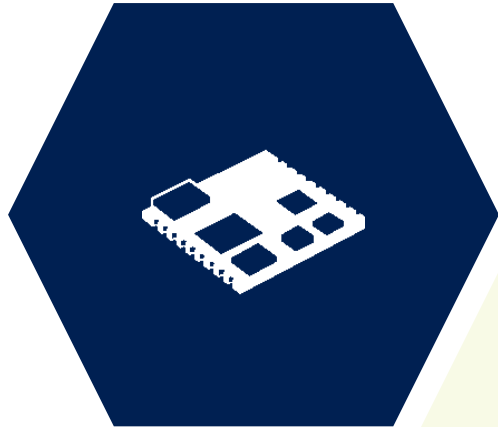
Process & analyze new data using trained NN



Form Factor HW

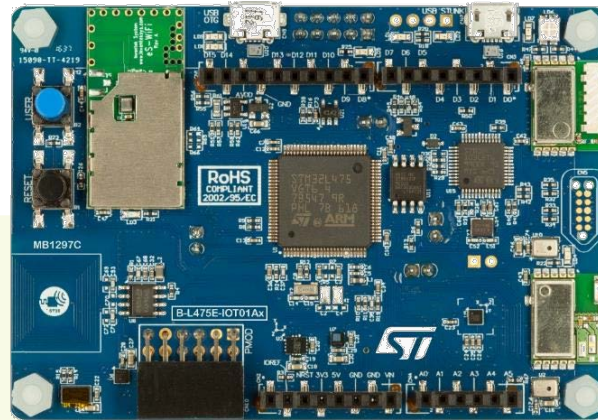
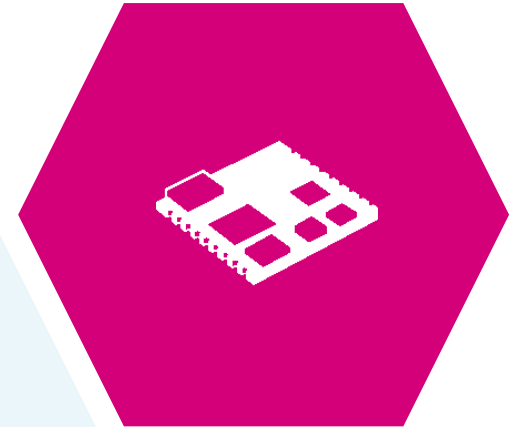
AI IoT Node for More Connectivity

Capture data



+  Sub-1GHz  Dynamic NFC Tag
 Wi-Fi

Process & analyze
new data using trained NN



More debug capabilities

- Integrated ST-Link/V2.1
- PMOD extension connector
- Arduino Uno extension connectors

<https://www.st.com/iotnode>





Food Recognition

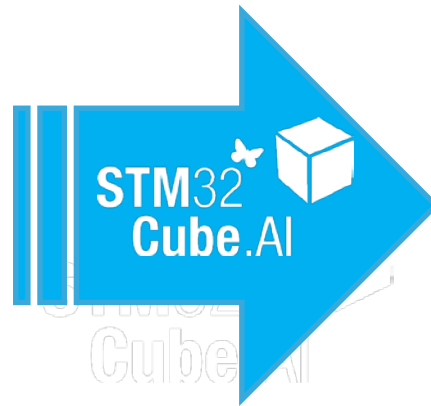
Fast Downsampling MobileNet Food Recognition on STM32H747 Dual-Core Discovery board

Neural Network

- FD-MobileNet topology from public paper applied to food
- Dataset

Implementation Details

- Uses Camera either in continuous or one shot mode
- Floating Point or mixed model Floating/Fix Point
- 18 food classes



STM32 Cube.AI NN

- Memory footprint: 205 KB RAM, 191 KB Flash

Performance on STM32H747

- 1 inference per image
- STM32H747 400 MHz Cortex-M7F
- Mix model Fix/Floating Point
 - 6.2 MHz / 150 ms per inference
 - Accuracy: 78.8%



Handwriting Character Recognition

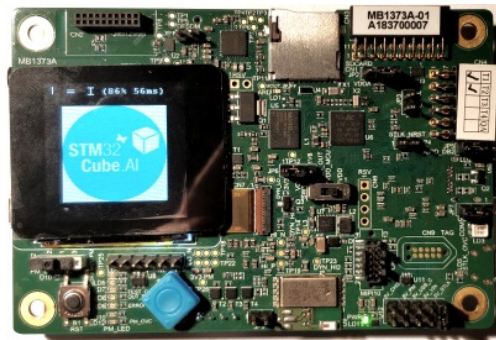
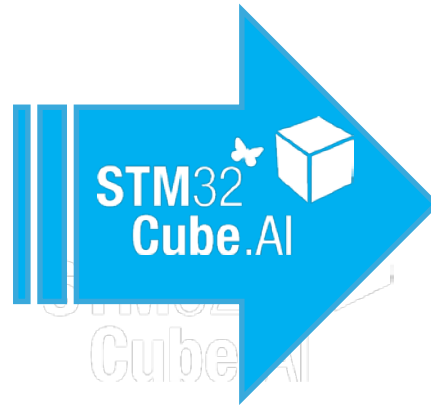
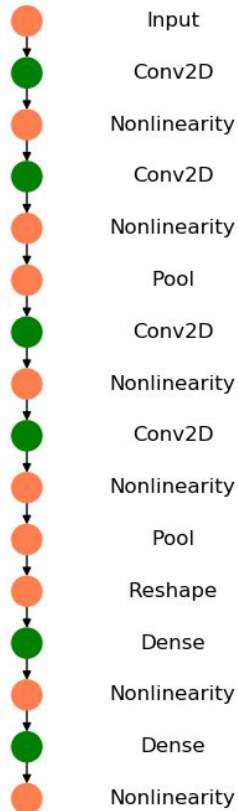
Demo: HCR on STM32L552 Discovery Board

Neural Network

- ST CNN
- EMNIST dataset (36 classes)

Implementation Details

- Exploits touch screen captured as image of size 32x32
- 36 classes: numbers and capital letters



STM32 Cube.AI NN

- Computational complexity 73k MACC
- Memory footprint: 26 KB RAM, 291 KB Flash

Performance on STM32L552

- 1 inference per image
- STM32L552 110 MHz Cortex-M33F
 - 6.2 MHz / 56 ms per inference
- STM32L496 80 MHz Cortex-M4F
 - 8 MHz / 100 ms per inference

